Received: 27 September 2020 | Revised: 26 October 2020 | Accepted: 7 November 2020

DOI: 10.1002/ima.22527

RESEARCH ARTICLE

WILEY

Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism

Tongxue Zhou[1,2,3] | **Stéphane Canu**[2,3] | **Su Ruan**[1,3]

[1]Université de Rouen Normandie, LITIS—QuantIF, Rouen, France

[2]INSA de Rouen, LITIS—Apprentissage, Rouen, France

[3]Normandie Univ, INSA Rouen, UNIROUEN, UNIHAVRE, LITIS, France

Correspondence
Su Ruan, Université de Rouen Normandie, LITIS—QuantIF, Rouen 76183, France.
Email: su.ruan@univ-rouen.fr

Abstract

The coronavirus disease (COVID-19) pandemic has led to a devastating effect on the global public health. Computed Tomography (CT) is an effective tool in the screening of COVID-19. It is of great importance to rapidly and accurately segment COVID-19 from CT to help diagnostic and patient monitoring. In this paper, we propose a U-Net based segmentation network using attention mechanism. As not all the features extracted from the encoders are useful for segmentation, we propose to incorporate an attention mechanism including a spatial attention module and a channel attention module, to a U-Net architecture to re-weight the feature representation spatially and channel-wise to capture rich contextual relationships for better feature representation. In addition, the focal Tversky loss is introduced to deal with small lesion segmentation. The experiment results, evaluated on a COVID-19 CT segmentation dataset where 473 CT slices are available, demonstrate the proposed method can achieve an accurate and rapid segmentation result on COVID-19. The method takes only 0.29 second to segment a single CT slice. The obtained Dice Score and Hausdorff Distance are 83.1% and 18.8, respectively.

KEYWORDS

attention mechanism, COVID-19, CT, deep learning, focal tversky loss, segmentation

1 | INTRODUCTION

In December 2019, a novel coronavirus, now designated as COVID-19 by the World Health Organization (WHO), was identified as the cause of an outbreak of acute respiratory illness.[1,2] The pandemic of COVID-19 is spreading all over the world and causes a devastating effect on the global public health. As a form of pneumonia, the infection causes inflammation in alveoli, which fills with fluid or pus, making the patient difficult to breathe.[3] Similar to other coronaviral pneumonia such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), COVID-19 can also lead to acute respiratory distress syndrome (ARDS).[4,5] In addition, the number of people infected by the virus is increasing rapidly. Up to August 11, 2020, 19 936 210 cases of COVID-19 have been reported in over 200 countries and territories, resulting in approximately 732 499 deaths,* while there is no efficient treatment at present.

Due to the fast progression and infectious ability of the disease, it is urgent to develop some tools to accurate diagnose and evaluate the disease. Although the real-time polymerase chain reaction (RT-PCR) assay of the sputum is considered as the gold standard for diagnosis, while it is time-consuming and has been reported to suffer from high false negative rates.[6,7] In clinical practice, Chest Computed tomography (CT), as a non-invasive imaging approach, can detect certain characteristic manifestations in the lung associated with COVID-19, for example, ground-glass opacities and consolidation are the most relative imaging features in pneumonia associated with SARS-CoV-2 infection. Therefore, Chest CT is considered as a low-cost, accurate, and efficient method diagnostic tool for early screening and diagnosis of COVID-19. It can be evaluated how severely the lungs are affected, and how the patient's disease is evolving, which is helpful in making treatment decisions.[8-12]

A number of artificial intelligence (AI) systems based on deep learning have been proposed and results have been shown to be quite promising in medical image analysis.[13-16] Compared to the traditional imaging workflow heavily relies on the human labors, AI enables more safe, accurate, and efficient imaging solutions. Recent AI-empowered applications in COVID-19 mainly include the dedicated imaging platform, the lung and infection region segmentation, the clinical assessment and diagnosis, as well as the pioneering basic and clinical research. Segmentation is an essential step in AI-based COVID-19 image processing and analysis for make a prediction of disease evolution. It delineates the regions of interest (ROIs), for example, lung, lobes, bronchopulmonary segments, and infected regions or lesions, in the chest X-ray or CT images for further assessment and quantification.[17] There are a number of researches related to COVID-19. For example, Zheng et al[18] proposed a weakly-supervised deep learning-based software system using 3D CT volumes to detect COVID-19. Goze et al[19] presented a system that utilizes 2D slice analysis and 3D volume analysis to achieve the detection of COVID-19. Jin et al[20] proposed an AI system for fast COVID-19 diagnosis, where a segmentation model is first used to obtain the lung lesion regions, and then the classification model is used to determine whether it is COVID-19-like for each lesion region. Li et al[8] developed a COVID-19 detection neural network (COVNet) to extract visual features from volumetric chest CT exams for distinguishing COVID-19 from Community Acquired Pneumonia (CAP). Chen et al[21] proposed to use Unet++[22] to extract valid areas and detect suspicious lesions in CT images.

U-net[23] is the most widely used encoder-decoder network architecture for medical image segmentation, since the encoder captures the low-level and high-level features, and the decoder combines the semantic features to construct the final result. However, not all features extracted from the encoder are useful for segmentation. Therefore, it is necessary to find an effective way to fuse features, we focus on the extraction of the most informative features for segmentation. Hu et al[24] introduced the Squeeze and Excitation (SE) block to improve the representational power of a network by modeling the interdependencies between the channels of its convolutional features. Roy et al[25] introduced to use both spatial and channel SE blocks (scSE), which concurrently recalibrates the feature representations spatially and channel-wise, and then combine them to obtain the final feature representation. Inspired by this work, we incorporate an attention mechanism including both spatial attention and channel one to our segmentation network to extract more informative feature representation to enhance the network performance.

In this paper, we propose a deep learning based segmentation with the attention mechanism. A preliminary conference version appeared at ISBI 2020,[26] which focused on the multi-model fusion issue. This journal version is a substantial extension, including (a) An automatic COVID-19 CT segmentation network. (b) A focal tversky loss function (different from the paper of ISBI) which is introduced to help to segment the small COVID-19 regions. (c) An attention mechanism including a spatial attention module and a channel attention module is introduced to capture rich contextual relationships for better feature representations.

The paper is organized as follows: Section 2 offers an overview of this work and details our model, Section 3 describes experimental setup, Section 4 presents the experimental results, Section 5 discusses the proposed method and Section 6 concludes this work.

2 | METHOD

2.1 | The proposed network architecture

Our network is mainly based on the U-Net architecture,[23] in which we integrate an attention mechanism, res_dil block and deep supervision. The encoder of the U-Net is used to obtain the feature representations. The feature representation at each layer are input into an attention mechanism, where they will be re-weighted along channel-wise and space-wise, and the most informative representations can be obtained, and finally they are projected by decoder to the label space to obtain the segmentation result. In the following, we will describe the main components of our model: encoder, decoder, and res_dil block, deep supervision and attention mechanism. The network architecture scheme is described in Figure 1.

2.2 | Encoder and decoder

The encoder is used to obtain the feature representations. It includes a convolutional block, a res_dil block followed by skip connection. In order to maintain the spatial

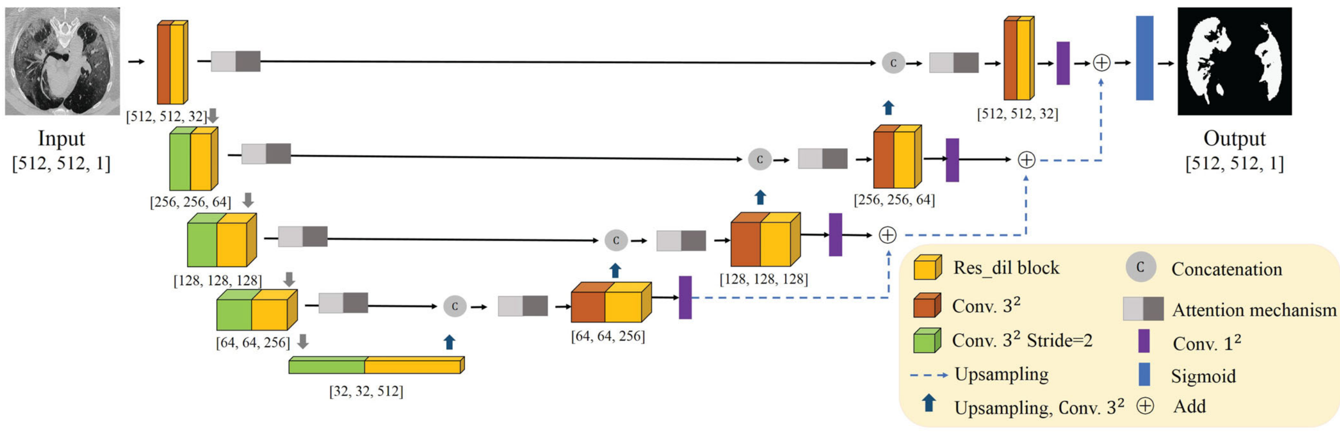

FIGURE 1 The architecture of the proposed network. The network takes a CT slice as input and directly outputs the COVID-19 region [Color figure can be viewed at wileyonlinelibrary.com]

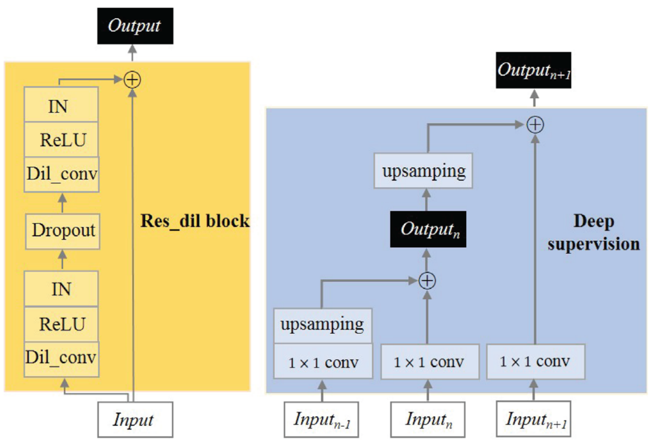

FIGURE 2 The architecture of our proposed Res dil block (left) and Deep supervision (right). IN refers instance normalization, Dil conv the dilated convolution (rate = 2, 4, respectively). We refer to the vertical depth as level, with higher levels being higher spatial resolution. In the deep supervision part, Input n refers the output of res dil block of the *n*th level in the decoder, Output n refers the segmentation result of the *n*th level in the decoder [Color figure can be viewed at wileyonlinelibrary.com]

information, we use a convolution with stride = 2 to replace pooling operation. It is likely to require different receptive field when segmenting different regions in an image. All convolutions are 3 × 3 and the number of filter is increased from 32 to 512. Each decoder level begins with up-sampling layer followed by a convolution to reduce the number of features by a factor of 2. Then the upsampled features are combined with the features from the corresponding level of the encoder part using concatenation. After the concatenation, we use the res_dil block to increase the receptive field. In addition, we employ deep supervision[27] for the segmentation decoder by integrating segmentation layers from different levels to form the final network output, shown in Figure 2.

2.3 | Res_dil block

It is likely to require different receptive field when segmenting different regions in an image. Since standard U-Net cannot get enough semantic features due to the limited receptive field, inspired by dilated convolution,[28] we proposed to use residual block with dilated convolutions on both encoder part and decoder part to obtain features at multiple scales, the architecture of res_dil is shown in Figure 2. The res_dil block can obtain more extensive local information to help retain information and fill details during training process.

To demonstrate that the proposed res_dil can enlarge the receptive field mathematically, we let $F: Z^2 \rightarrow R$ be a discrete function, $\Omega_r = [-r, r]^2 \in Z^2$and let $k: \Omega_r \rightarrow R$ be a discrete filter size $(2r+1)^2$. The discrete convolution operator $\star$ can be described as follows:

$$(F \star k) = \sum\nolimits_{m=-r}^{r} \sum\nolimits_{n=-r}^{r} F(x-m, y-n)k(m,n) \quad (1)$$

Let l be a dilation factor and the l-dilated convolution operation $\star_l$ can be defined as:

$$(F \star_l k) = \sum\nolimits_{m=-r}^{r} \sum\nolimits_{n=-r}^{r} F(x-lm, y-ln)k(m,n) \quad (2)$$

We assume $F_0, F_1, \ldots, F_{n-1}: Z^2 \rightarrow R$ are a discrete functions, and $k_0, k_1, \ldots, k_{n-2}: Z^2 \rightarrow R$ are discrete 3 × 3 filters. In addition, we apply the filters with exponentially increasing dilation factors, such as $2^0, 2^1, \ldots, 2^{n-2}$. Then, the discrete function F_{i+1} can be described as:

$$F_{i+1} = F_i \star_{2i} k_i, i = 0, 1, ..., n-2 \tag{3}$$

According to the definition of receptive field, the receptive field size of each element in F_{i+1} is $(2^{i+2} - 1) \times (2^{i+2} - 1)$, which is a square of exponentially increasing size. So we can obtain a 15×15 receptive field by applying our proposed res_dil block with the dilation factor 2 and 4, respectively, while the classical convolution can only obtain 7×7 receptive field, see Figure 3.

Since there exists many small regions of interests (ROIs) in a COVID-19 CT image, therefore, increasing the receptive field of the feature representation is essential and it can help the network to extract more contextual information to achieve a better segmentation result.

2.4 | Attention mechanism

In U-net shaped network, not all the features obtained by the encoder are effective for segmentation. In addition, not only the different channels (filters) have various contributions but also different spatial location in each channel can give different weights on feature representation for segmentation. To this end, we introduced a "scSE based" attention mechanism in both encoder and decoder to take into account the most informative feature representations along channel-wise and spatial-wise for segmentation, the architecture is described in Figure 4.

The individual feature representations from each channel are first concatenated as the input representation $Z = [z_1, z_2, ..., z_n]$, $Z_k \in R^{H \times W}$, n is the number of channel in each layer. To simplify the description, we take $n = 32$.

In the channel attention module, a global average pooling is first performed to produce a tensor $g \in R^{1 \times 1 \times 32}$, which represents the global spatial information of the representation, with its k^{th} element

$$g_k = \frac{1}{H \times W} \sum_i^H \sum_j^W Z_k(i,j) \tag{4}$$

Then two fully-connected layers are applied to encode the channel-wise dependencies, $\hat{g} = W_1(\delta(W_2 g))$, with $W_1 \in R^{32 \times 16}, W_2 \in R^{16 \times 32}$, being weights of two fully-connected layers and the ReLU operator $\delta(\cdot)$. $\hat{g}$ is then passed through the sigmoid layer to obtain the channel-wise weights, which will be applied to the input representation Z through multiplication to achieve the channel-wise representation Z_c, the $\sigma(\hat{g_k})$ indicates the importance of the i channel of the representation:

$$Z_c = [\sigma(\hat{g_1}) z_1, \sigma(\hat{g_2}) z_2, ..., \sigma(\hat{g_{32}}) z_{32}] \tag{5}$$

In the spatial attention module, the representation can be considered as $Z = [z^{1,1}, z^{1,2}, ..., z^{i,j}, ..., z^{H,W}]$, $Z^{i,j} \in R^{1 \times 1 \times 32}$, $i \in 1, 2, ..., H, j \in 1, 2, ..., W$ and then a convolution operation $q = W_s \star Z, q \in R^{H \times W}$ with weight $W_s \in R^{1 \times 1 \times 32 \times 1}$, is used to squeeze the spatial domain, and to produce a projection tensor, which represents the linearly combined representation for all channels for a spatial location. The tensor is finally passed through a sigmoid layer to obtain the space-wise weights and to

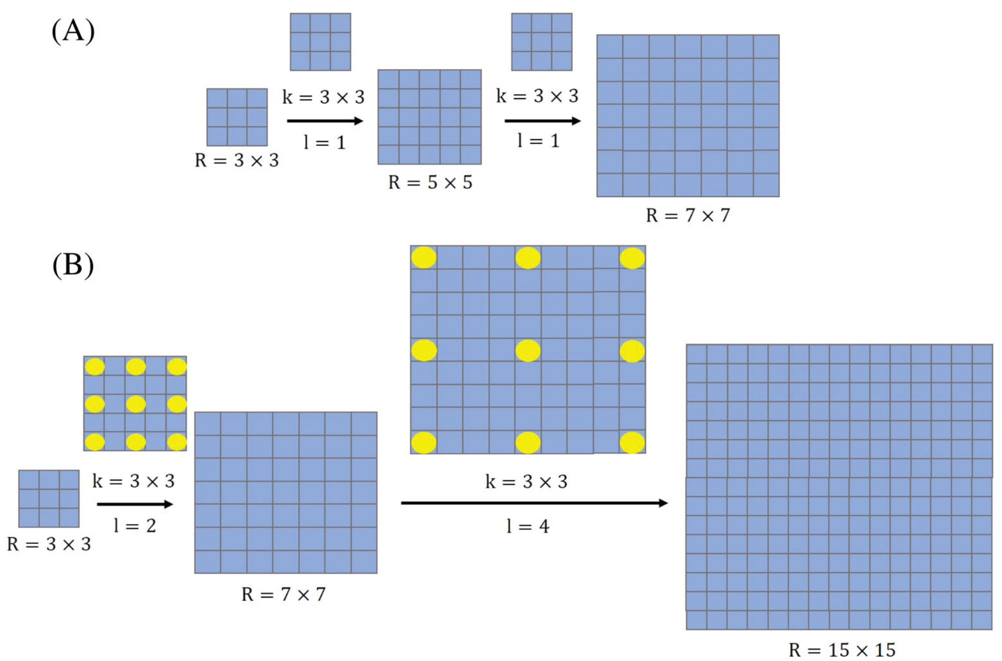

FIGURE 3 The illustration of receptive field, R denotes the receptive field, k denotes the convolution kernel size, and l denotes the dilated factor. A, A convolution network which consists of two $k = 3 \times 3$ and $l = 1$, 1 convolutional layers. B, A convolution network which consists of two $k = 3 \times 3$ and $l = 2$, 4 dilated convolutional layers [Color figure can be viewed at wileyonlinelibrary.com]

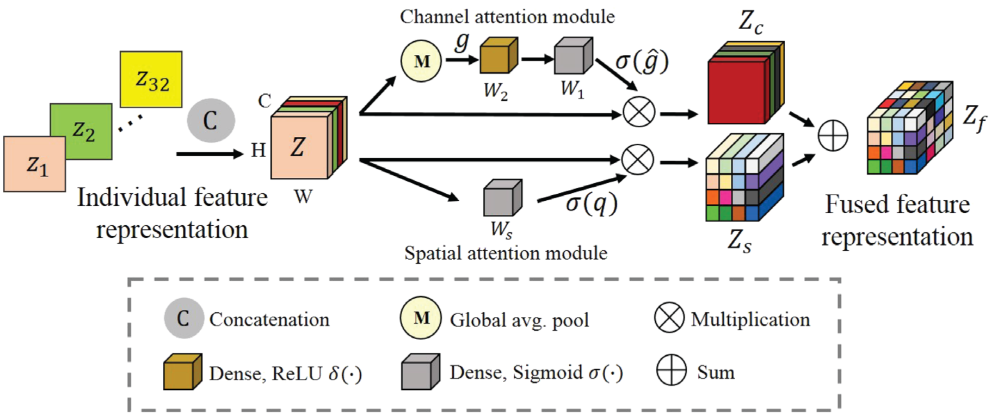

FIGURE 4 The architecture of attention mechanism. The individual feature representations (z_1, z_2, ..., z_{32}) are first concatenated as Z, and then they are recalibrated spatially and channel-wise to achieve the Z_s and Z_c, final they are added to obtain the rich fused feature representation Z_f [Color figure can be viewed at wileyonlinelibrary.com]

achieve the spatial-wise representation Z_s, the $\sigma(q_{i,j})$ that indicates the importance of the spatial information (i, j) of the representation:

$$Z_s = \left[\sigma(q_{1,1})z^{1,1}, \ldots, \sigma\left(q_{i,j}\right)z^{i,j}, \ldots, \sigma(q_{H,W})z^{H,W}\right] \tag{6}$$

The fused feature representation is obtained by adding the channel-wise representation and space-wise representation:

$$Z_f = Z_c + Z_s \tag{7}$$

The attention mechanism can be directly adapted to any feature representation problem, and it encourages the network to capture rich contextual relationships for better feature representations.

2.5 | Loss function

In the medical community, the Dice Score Coefficient (DSC), defined in (8), is the most widespread metric to measure the overlap ratio of the segmented region and the ground truth, and it is widely used to evaluate segmentation performance. Dice Loss (DL) in (9) is defined as a minimization of the overlap between the prediction and ground truth.

$$\mathrm{DSC}_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^{N} (p_{ic} + g_{ic}) + \varepsilon} \tag{8}$$

$$\mathrm{DL}_c = \sum_c (1 - \mathrm{DSC}_c) \tag{9}$$

where N is the number of pixels in the image, c is the set of the classes, p_{ic} is the probability that pixel i is of the lesion class c, the same is true for g_{ic}, ε is a small constant to avoid dividing by 0.

One of the limitation of Dice Loss is that it penalizes false positive (FP) and false negative (FN) equally, which results in segmentation maps with high precision but low recall. This is particularly true for highly imbalanced dataset and small regions of interests (ROI) such as COVID-19 lesions. Experimental results show that FN needs to be weighted higher than FP to improve recall rate. Tversky similarity index[29] is a generalization of the DSC which allows for flexibility in balancing FP and FN:

$$\mathrm{TI}_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{i'c} g_{ic} + \beta \sum_{i=1}^{N} p_{ic} g_{i'c} + \varepsilon} \tag{10}$$

where N is the number of pixels in the image, c is the set of the classes, p_{ic} is the probability that pixel i is of the lesion class c and $p_{i'c}$ is the probability that pixel i is of the non-lesion class $'c$, the same is true for g_{ic} and $g_{i'c}$, ε is a small constant to avoid dividing by 0. When $\alpha = \beta = 0.5$, TI_c is the same as DL_c, in our work, $\alpha = 0.7, \beta = 0.3$.

Another issue with the DL is that it struggles to segment small ROIs as they do not contribute to the loss significantly. To address this, Abraham et al[30] proposed the Focal Tversky Loss function (FTL).

$$\mathrm{FTL}_c = \sum_c (1 - \mathrm{TI}_c)^{\frac{1}{\gamma}} \tag{11}$$

where γ varies in the range [1, 3], in our work, $\gamma = \frac{4}{3}$. In practice, if a pixel is misclassified with a high Tversky index, the FTL is unaffected. However, if the Tversky index is small and the pixel is misclassified, the FTL will decrease significantly. To this end, we used FTL to train the network to help segment the small COVID-19 regions.

3 | EXPERIMENTAL SETUP

3.1 | Dataset and preprocessing

The two datasets used in the experiments come from Italian Society of Medical and Interventional Radiology: COVID-19 CT segmentation dataset.† Dataset-1 includes 100 axial CT images from 60 patients with Covid-19. The images have been resized, grayscaled, and compiled into a single NIFTI-file. The image size is 512×512 pixels. The images have been segmented by a radiologist using three labels: ground-glass, consolidation and pleural effusion. Dataset-2 includes 9 volumes, total 829 slices, where 373 slices have been evaluated and segmented by a radiologist as COVID-19 cases. We resize these images from 630×630 pixels to 512×512 pixels same as Dataset-1. And an intensity normalization is applied to both datasets. Since there are severe data imbalance in the dataset. For example, in Dataset-1, only 25 slices have pleural effusion, which is the smallest region among all the COVID-19 lesion regions (see the green region in Figure 5). In Dataset-2, only 233 slices have consolidation, which takes up a small amount of pixels in the image (see the yellow region in Figure 5). We take all the lesion labels as a COVID-19 lesion. Because of the small number of data in both two datasets, we combine the two datasets as our final training dataset, finally, 473 CT slices are used to train our model. Here, we give some example images of the COVID-19 CT segmentation dataset in Figure 5.

3.2 | Implementation details

Our network is implemented in Keras with a single Nvidia GPU Quadro P5000 (16G). The network is trained by focal tversky loss and is optimized using the Adam optimizer, the initial learning rate $= 5e-5$ with a decreasing learning rate factor 0.5 with patience of 10 epochs. Early stopping is employed to avoid overfitting if the validation loss is not improved over 50 epochs. We randomly split the dataset into 80% training and 20% testing.

3.3 | Evaluation metrics

Segmentation accuracy determines the eventual success or failure of segmentation procedures. To measure the segmentation performance of the proposed methods, two evaluation metrics: Dice and Hausdorff Distance are used to obtain quantitative measurements of the segmentation accuracy.

3.3.1 | Dice Score

It is designed to evaluate the overlap rate of prediction results and ground truth. Dice Score ranges from 0 to 1, and the better predict result will have a larger Dice Score value.

$$\text{Dice Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{12}$$

where TP represents the number of true positive voxels, FP represents the number of false positive voxels, and FN represents the number of false negative voxels.

3.3.2 | Hausdorff Distance (HD)

It is computed between boundaries of the prediction results and ground-truth, it is an indicator of the largest segmentation error. The better predict result will have a smaller HD value.

$$\text{HD} = \max\{r \in \partial R d_m(s,r), s \in \partial S d_m(r,s)\} \tag{13}$$

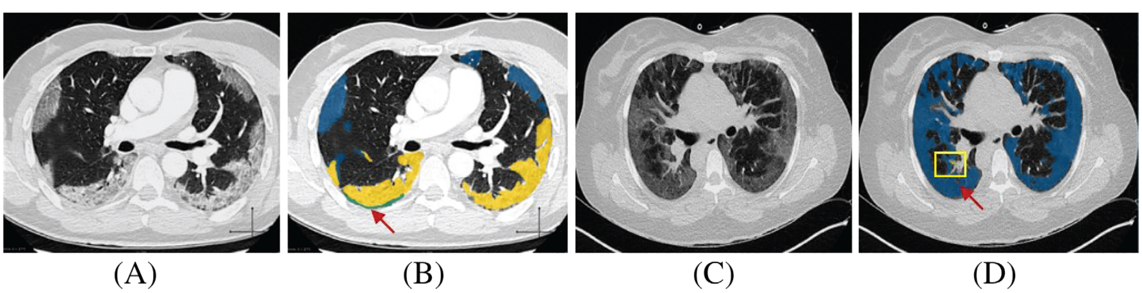

FIGURE 5 Example images of the COVID-19 CT segmentation dataset. A and C, CT image from Dataset-1 and Dataset-2; B and D, ground truth of panels A, C, respectively, ground-glass is shown in blue, consolidation is shown in yellow and pleural effusion is shown in green [Color figure can be viewed at wileyonlinelibrary.com]

where ∂S and ∂R are the sets of lesion border pixels for the predicted and the real annotations, and $d_m(v, v)$ is the minimum of the Euclidean distances between a voxel v and voxels in a set v.

4 | EXPERIMENT RESULTS

In this section, we conduct extensive comparative experiments including quantitative analysis and qualitative analysis to demonstrate the effectiveness of our proposed method. In Section 4.1.1, we first perform an ablation experiment to see the importance of our proposed res_dil block. Then in Section 4.1.2, we analyze the performance of our proposed method trained by Focal Tversky Loss function. In Section 4.1.3, we validate the contribution of proposed attention mechanism based fusion block. In Section 4.1.4, we compare our method with the state-of-the-art methods. In Section 4.2.1 the qualitative experiments of our method and the qualitative comparison experiments in Section 4.2.2 with the state-of-the-art methods are carried out to further demonstrate the contribution of our proposed method.

4.1 | Quantitative analysis

In this section, we conduct several experiments to validate the performance of each key component of our method, including the res_dil block, focal tversky loss and attention mechanism. Furthermore, we compare our method with the state-of-the-art methods.

4.1.1 | Performance analysis of res_dil block

To assess the performance of our method, and to analyze the impact of the proposed components of our network, we first did an ablation study, the results are shown in Table 1. With regard to the proposed res_dil block, we can observe that the proposed res_dil block can boost the "Backbone + DL" with the improvement of 0.12% and 3.70% in the terms of Dice Score and Hausdorff Distance. Also we can see an improvement of 0.25% and 46.48% in the terms of Dice Score and Hausdorff Distance compared to "Backbone + FTL". We explain that the larger receptive region obtained from res_dil block can help the network to capture more rich feature information in order to achieve a better segmentation result. The results clearly show that the proposed res_dil block is necessary for boosting the segmentation performance.

TABLE 1 Comparison of different methods on COVID-19 CT segmentation dataset, bold results show the best scores

Methods	Dice score (%)	Hausdorff Distance (mm)
Backbone + DL	80.7	43.2
Backbone + FTL	80.9	35.5
Backbone + Res_dil + DL	80.8	41.6
Backbone + Res_dil + FTL	81.1	19.01
Backbone + Attention + DL	82.2	34.2
Backbone + Attention + FTL	82.4	32.3
Backbone + Res_dil + Attention + DL	82.6	30.7
Backbone + Res_dil + Attention + FTL	**83.1**	**18.8**

4.1.2 | Performance analysis of FTL

We also demonstrate the effectiveness of applying the Focal Tversky Loss function (FTL). From Table 1, we can observe the proposed method (Backbone + Res_dil + Attention) trained with DL achieves Dice Score, and HD of 82.6% and 30.7, respectively. However, using FTL can aide the network to focus more on the false negative voxels, which increases 0.61% of Dice Score and 38.76% of Hausdorff Distance. This suggests that applying the FTL to train our model can achieve the better results.

4.1.3 | Performance analysis of attention mechanism

To investigate the contribution of proposed attention mechanism, we also did an another ablation experiment in Table 1. We can observe that integrating the attention mechanism to the "Backbone + DL" method can boost the performance, since we can see an increase of 1.86% of Dice Score and 20.83% of Hausdorff Distance, and also an improvement of 1.85% of Dice Score compared to "Backbone + FTL." The main reason is that the attention mechanism can help to emphasis on the most important feature representation for segmentation. In addition, the proposed network trained by FTL combines the benefits of attention mechanism can obtain the best results with Dice = 83.1% and HD = 18.8, which has an improvement of 2.97% and 56.48% in the terms of Dice Score and Hausdorff Distance compared to the "Backbone + DL."

To further demonstrate the contribution of attention mechanism, we select three examples to visualize the feature maps in Figure 6. The first column shows the input CT image, the second column shows the ground truth, the third and fourth columns show the feature maps before and after using the attention mechanism. From the results, we can observe that without using the attention mechanism, the network cannot capture all the interested lesion region or just capture a part of interested segmentation regions, however, applying the attention mechanism can help the network learn more useful feature information for the final segmentation, and we can see the clearer ROIs of the segmentation. The visualization results further validate the effectiveness of the proposed attention mechanism.

4.1.4 | Comparison with the state-of-the-art methods

We compare our method with the state-of-the-art methods including Unet,[23] Unet++,[22] Attention-Unet,[31] the quantitative results are shown in Table 2. As we can see, the classic U-Net can achieve the Dice Score and Hausdorff Distance of 82.5% and 23.4, respectively. While the U-Net++ obtains a better results thanks to the nested connection between encoder and decoder, which reduces the semantic gap between the feature maps of the encoder and decoder sub-networks. However, the improvement is still not impressive. Compared to these two state-of-the-art approaches, the attention U-Net has a worse result even if the attention gates are used. However, our proposed method outperforms all the methods by a large margin, which achieves the best segmentation results across all the evaluation metrics. We attribute the improvement to the proposed components including res_dil block, attention mechanism, which can help the network capture more useful feature information to enhance the segmentation. Also, the FTL can aide the network to achieve a better performance on the small ROI.

4.2 | Qualitative analysis

In order to demonstrate the effectiveness of our model, we randomly select several examples on COVID-19 CT segmentation dataset and visualize the results in Figures 7 and 8.

TABLE 2 Comparison with the state-of-the-art methods on COVID-19 CT segmentation dataset, bold results show the best scores

Methods	Dice score (%)	Hausdorff distance (mm)
Attention-U-Net[31]	75.5	41.3
U-Net (MICCAI'15)[23]	82.5	23.4
U-Net ++(TMI'19)[22]	82.6	22.2
Ours	**83.1**	**18.8**

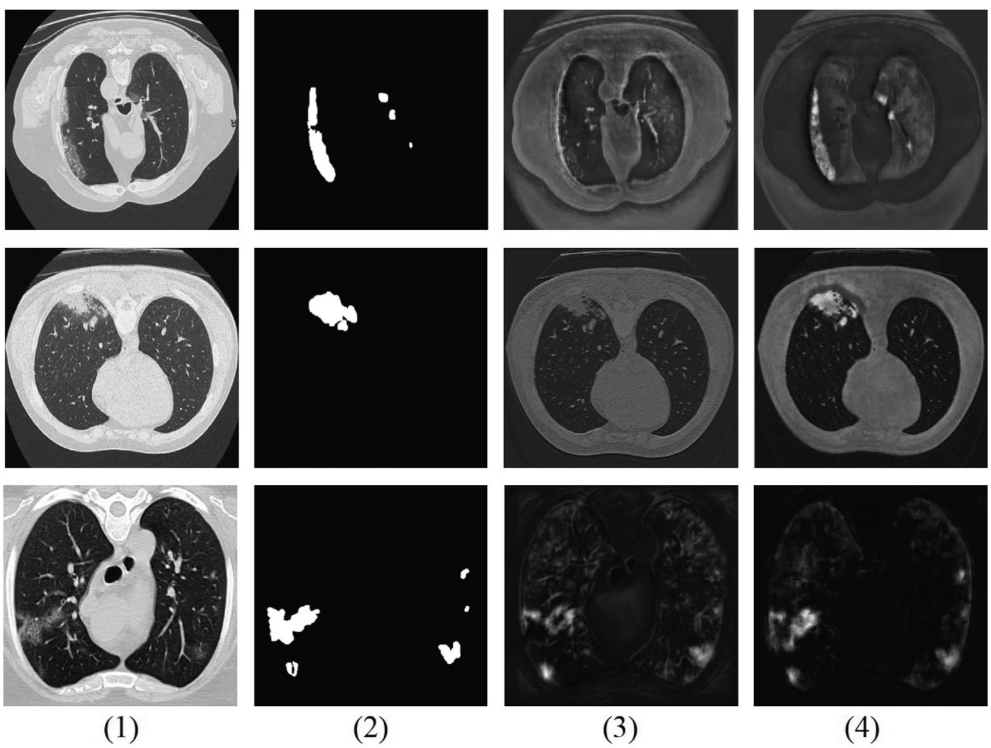

FIGURE 6 Visualization of proposed attention mechanism. The rows show the examples, the column (1) input CT image, (2) ground truth, (3) before using attention mechanism, (4) after using attention mechanism

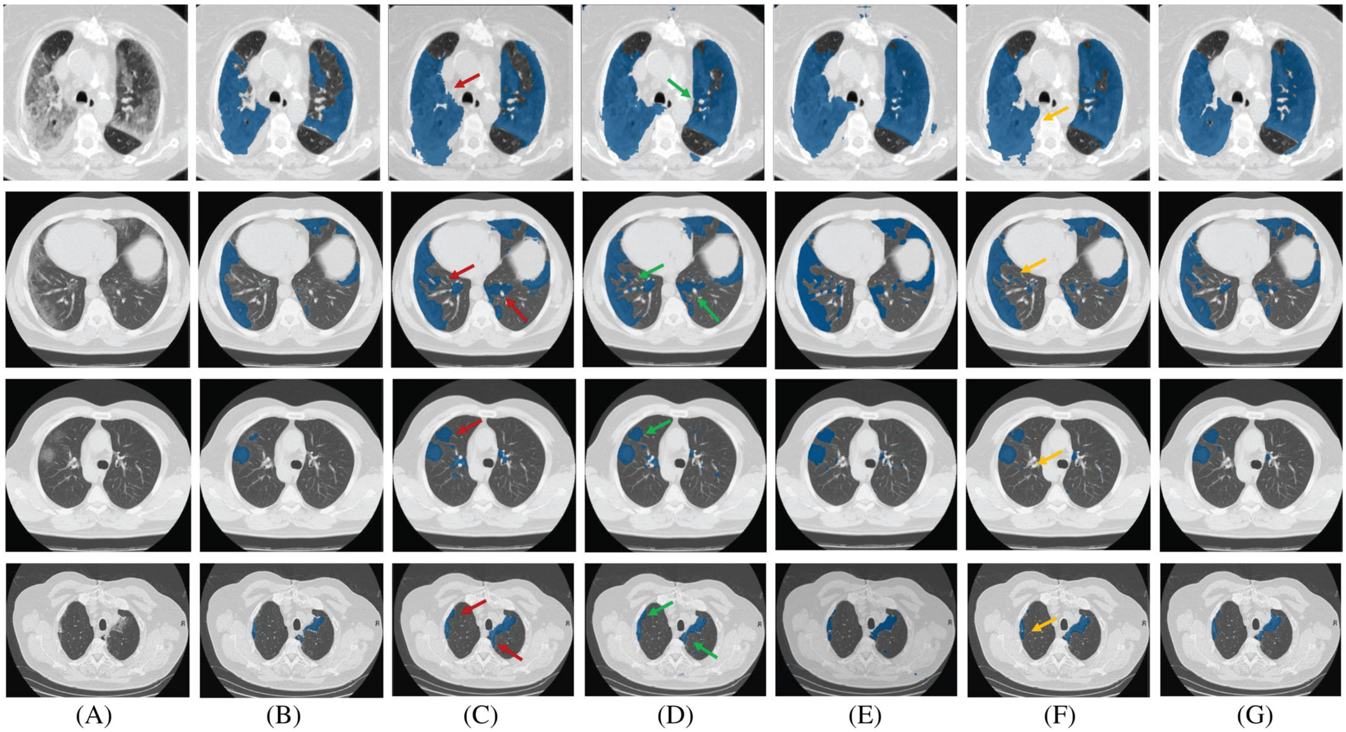

FIGURE 7 Segmentation results of some examples on COVID-19 CT dataset. The first two examples are with many COVID-19 lesion regions, the last two examples are with few COVID-19 regions. A, CT image; B, Backbone + DL; C, Backbone + Res dil + DL; D, Backbone + Attention + DL; E, Backbone + Res dil + Attention + DL, F, Backbone + Res dil + Attention + FTL; G, Ground truth, red arrow emphasizes the improvement of using res dil block (from B to C), green arrow emphasizes the improvement of applying attention mechanism (from B to D), yellow arrow emphasizes the improvement of applying FTL (from E to F) [Color figure can be viewed at wileyonlinelibrary.com]

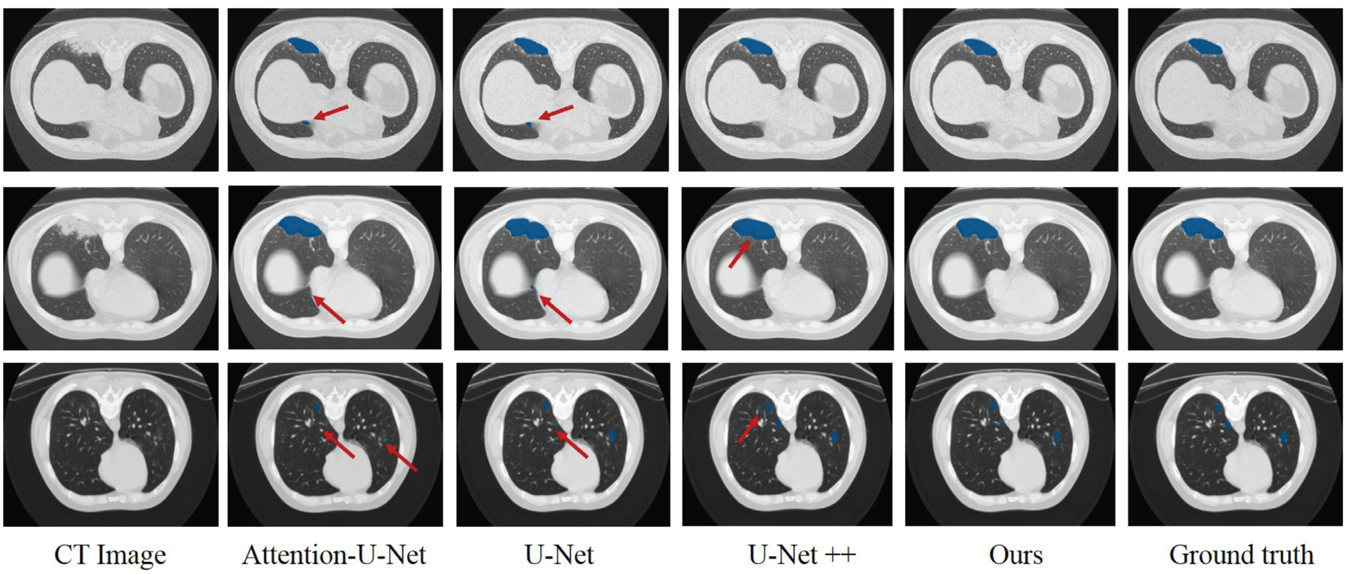

FIGURE 8 Segmentation results of some examples between different methods on COVID-19 CT dataset. Red arrow emphasizes the mis-segmentated regions of each method [Color figure can be viewed at wileyonlinelibrary.com]

4.2.1 | Visualization results of our proposed method

From Figure 7, we can observe that the backbone trained by DL could give a rough segmentation result, while it fails to segment many small lesion regions. With the application of res_dil block and attention mechanism, it can be seen that the proposed res_dil block enhance the segmentation results benefitting from the larger receptive field. In addition, the attention mechanism can help to capture more rich feature information to further refine the segmentation result. Compared to panel E, the proposed network trained by FTL (F) can achieve the result closest to the ground truth. The obtained results have demonstrated that leveraging the res_dil block, the attention mechanism and the FTL can generally enhance the COVID-19 segmentation performance.

4.2.2 | Comparison with the state-of-the-art methods

We also visualize the comparison results with the state-of-art methods in Figure 8. From the results, we can observe that the proposed model can detect the lesion regions effectively. Also, the segmentation results are close to the ground truth. On the contrary, the Attention-U-Net and U-Net give unsatisfied results, from the first two examples, we can see some non-target regions are detected. And in the third example, some target regions cannot be segmented. However, U-Net ++ can achieve a better result for the first example, but still not promising in the last two examples. As can be observed, compared with these three methods, our proposed method yields more accurate segmentation results, we explain the success of our method is due to all the proposed components in the network.

5 | DISCUSSION

Due to the fast progression and infectious ability of the COVID-19, it is necessary to develop some tools to accurate diagnose and evaluate the disease. Recently, deep learning based methods have shown promising segmentation performance. To this end, we presented an automatic COVID-19 CT segmentation network. The network is based on the U-Net architecture and we integrated an attention mechanism, res_dil block and FTL loss to the network. The extensive ablation experiments in Section 4.1 demonstrate the effectiveness of each proposed component, and our method can achieve the best results when all the components are integrated together. Furthermore, to prove the effectiveness of proposed attention mechanism, we visualize the feature maps before and after using the attention mechanism. We can observe that the ROI of segmentation is clearer after applying the attention mechanism. We attribute it to the spatial and channel attention modules, which aide the network to extract rich contextual feature representation. In addition, we also compare our method with the state-of-the-art methods, and the quantitative and qualitative comparison results further prove the advantage of our proposed method.

The advantages of our proposed network architecture: (a) The experiment results evaluated on the two metrics (Dice Score and Hausdorff Distance) demonstrate that our proposed method can give an impressing segmentation result. (b) The comparison results with the other state-of-the-art approached demonstrate the contribution of our method. (c) The architecture is an end-to-end deep leaning approach and fully automatic without any user interventions. (d) The proposed attention based fusion block can be generalized to other multi-modal segmentation task.

However, our work has some limitations that inspire future directions. (a) The study is limited by the small dataset. Therefore, in the future, we would like to use a larger training dataset or apply the data augmentation techniques to achieve more competitive results. (b) The network is designed to segment the single label, we plan to apply our method to other multi-class segmentation tasks and compared with other related methods. (c) The proposed method is evaluated for the public COVID-19 segmentation dataset, in the future, we plan to validate our method on other segmentation tasks (eg, medical organ segmentation or non-medical segmentation.)

6 | CONCLUSION

In this paper, we have presented a U-Net based network using attention mechanism for COVID-19 segmentation. Since most current segmentation networks are trained with Dice loss, which penalize the false negative voxels and false positive voxels equally. To this end, we applied the focal tversky loss to train the model to improve the small ROI segmentation performance. IN addition, the res_dil block and the attention mechanism are used in each layer to capture rich contextual relationships for better feature representations. We evaluated our proposed network on COVID-19 CT segmentation datasets and compared with the state-of-the-art approaches, the experiment results demonstrate thesuperior performance of our method.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

AUTHOR CONTRIBUTIONS

Tongxue Zhou designed the model, analyzed the data, carried out the implementation, and wrote the manuscript. Stéphane Canu and Su Ruan provided suggestions for experiment and revised the manuscript critically.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in COVID-19 CT segmentation dataset at http://medicalsegmentation.com/covid19/, footnote [2].

ORCID

Tongxue Zhou https://orcid.org/0000-0003-3110-4884
Su Ruan https://orcid.org/0000-0001-8785-6917

ENDNOTES

* https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

† http://medicalsegmentation.com/covid19/.

REFERENCES

1. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: a modelling study. *The Lancet*. 2020;395(10225):689-697.
2. W. H. Organization. *Who Director-general's Opening Remarks at the Media Briefing on Covid-19-11*. Geneva, Switzerland; March 2020.
3. Shi F, Xia L, Shan F, et al. Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. 2020. *arXiv Preprint arXiv*.
4. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497-506.
5. Li H, Liu S-M, Yu X-H, Tang S-L, Tang C-K. Coronavirus disease 2019 (covid-19): current status and future perspective. *Int J Antimicrob Agents*. 2020;105951.
6. Shan+ F, Gao+ Y, Wang J, et al. Lung infection quantification of covid-19 in ct images with deep learning. *arXiv Preprint arXiv* 2020.
7. Liang T. *Handbook of Covid-19 Prevention and Treatment*. Zhejiang: Zhejiang University School of Medicine; 2020.
8. Li L, Qin L, Xu Z, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*. 2020;296:200905.
9. Pan F, Ye T, Sun P, et al. Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. *Radiology*. 2020;295:200370.
10. Ng M-Y, Lee EY, Yang J, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiol Cardiothorac Imaging*. 2020;2(1):e200034.
11. Lei J, Li J, Li X, Qi X. Ct imaging of the 2019 novel coronavirus (2019-ncov) pneumonia. *Radiology*. 2020;295(1):18-18.
12. Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest ct manifestations of new coronavirus disease 2019 (covid-19): a pictorial review. *Eur Radiol*. 2020;30:1-9.
13. Zhou T, Canu S, Vera P, Ruan S. Brain tumor segmentation with missing modalities via latent multi-source correlation representation. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., eds. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*. Cham: Springer International Publishing; 2020:533-541.
14. Zheng L, Wang G, Zhang F, Zhao Q, Dai C, Yousefi N. Breast cancer diagnosis based on a new improved elman neural network optimized by meta-heuristics. *Int J Imaging Syst Technol*. 2020;30(3):513-526.
15. Peng S, Chen W, Sun J, Liu B. Multi-scale 3d u-nets: an approach to automatic segmentation of brain tumor. *Int J Imaging Syst Technol*. 2020;30(1):5-17.
16. Hu H, Guan Q, Chen S, Ji Z, Lin Y. Detection and recognition for life state of cell cancer using two-stage cascade cnns. *IEEE/ACM Trans Comput Biol Bioinform*. 2020;17(3):887-898.
17. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE Reviews in Biomedical Engineering*. 2020; 13:1-1.
18. Zheng C, Deng X, Fu Q, et al. Deep learning-based detection for covid-19 from chest ct using weak label. *medRxiv*. 2020.
19. Gozes O, Frid-Adar M, Greenspan H, et al. Rapid ai development cycle for the coronavirus (covid-19) pandemic: initial results for automated detection & patient monitoring using deep learning ct image analysis. *arXiv Preprint arXiv*. 2020.
20. Jin S, Wang B, Xu H, et al. Ai-assisted ct imaging analysis for covid-19 screening: building and deploying a medical ai system in four weeks. *medRxiv*. 2020.
21. Chen J, Wu L, Zhang J, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. *medRxiv*. 2020.
22. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Switzerland: Springer; 2018:3-11.
23. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Switzerland: Springer; 2015:234-241.
24. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. US: IEEE; 2018:7132-7141.
25. Roy AG, Navab N, Wachinger C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Switzerland: Springer; 2018: 421-429.
26. Zhou T, Ruan S, Guo Y, Canu S. A multi-modality fusion network based on attention mechanism for brain tumor segmentation. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE*. US: IEEE; 2020:377-380.

27. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. *International MICCAI Brainlesion Workshop*. Switzerland: Springer; 2017:287-297.
28. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv Preprint arXiv*. 2015.
29. Tversky A. Features of similarity. *Psychol Rev*. 1977;84(4): 327-352.
30. Abraham N, Khan NM. A novel focal tversky loss function with improved attention u-net for lesion segmentation. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. US: IEEE; 2019:683-687.
31. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: learning where to look for the pancreas. *arXiv Preprint arXiv*. 2018.

How to cite this article: Zhou T, Canu S, Ruan S. Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. *Int J Imaging Syst Technol*. 2021;31: 16–27. https://doi.org/10.1002/ima.22527